

Book Reviews

Developing Bioinformatics Computer Skills. By Cynthia Gibas and Per Jambeck. Sebastopol, California: O'Reilly & Associates, 2001, 442 pp., \$34.95.

While the true definition of bioinformatics is elusive and consistent only in its variability, most would agree that the field demands the application of computer-driven data analysis and programming techniques to questions in molecular biology. As such, the area has one of the tallest entry requirements of any academic field; ideally, the student should be intricately familiar not only with biochemistry, molecular biology, biophysics, genetics, and laboratory techniques, but also computer programming, database design and the various strengths and limitations of computational approaches to data crunching. The bioinformatics scientist can be called upon to produce meaningful analyses of nearly any aspect of data gleaned from a number of genomics and proteomics approaches, as well as teasing functional clues out of vast online databank searches, and as such must be aware of the caveats and benefits of each technique involved. Such knowledge takes years to acquire, and becoming a competent bioinformatics scientist is an understandably challenging process. Nonetheless, there exists a need for biologists with a basic grounding in computational analysis of scientific problems, or perhaps also expert programmers primed in the broad generalities of molecular biology, and it is here that this book finds its audience. While presupposing no knowledge of either field, the authors lay

out the groundwork of both basic molecular biology and Unix-driven bioinformatics techniques. This is done effectively, though at the expense of continuity: the book's four main sections feel at times like several primers thrown together.

The introductory section, "Biology in the Computer Age," provides a good overview of the state of the field and the need for computational biology to manage vast and rapidly expanding repositories of data in an accessible, easily digested style. The second section, spanning nearly ninety pages, is completely concerned with setting up and using a Unix operating system. I found this section far less useful than the others; a better approach would be to list some available Unix/Linux platforms, the relative merits of each, and send the reader on his or her way to purchasing an O'Reilly Unix book, which they will more than likely have to do anyway. The third section, "Tools for Bioinformatics," gives an excellent overview of the various ways in which biological data can be analyzed, how the tools work, and things to keep in mind while using them. Of particular interest is the explanation of the NCBI BLAST protocol, which, in my view, should be highlighted and moved to the beginning of the book, since a basic understanding of this program's output is both necessary and very rare among laboratory research scientists attempting to interpret their database query results. The fourth and final section briefly explores databases and visualization, with much the same results as the third.

O'Reilly is a seasoned publishing house for computer reference guides, maintaining a consistent style in their titles. The sister book to this one, *Beginning Perl for Bioinformatics*, makes good use of this format in imparting the basics of Perl with a bio-scientific twist. *Developing Bioinformatics Computer Skills*, however, is extremely light on code and comparatively heavy on "how-to" sections that are likely to become quickly dated, and much of the text feels as though explanations of biological science have been pigeonholed into a format better suited for explaining C++ syntax. The presentation of simple, short explanations (complete with gimmicky "Caution" notes) on such topics as "Planning Array Experiments" is at once amusing and misleading, since the likelihood of readers setting up such experiments based in whole or in part on these paragraphs is vanishingly small.

The target audience of this book is listed as both students "learning computational approaches to biology for the first time" and experienced researchers "just starting to use computers" to handle their data. To me, bioinformatics has two general tiers of expertise: laboratory researchers exhaustively versed in biology who understand the basics of similarity searching, alignments and web-accessible computational services like transmembrane helix prediction; and bioinformaticians who code custom-built programs to solve novel or specialized problems in science. The requirements of each are considerably different, and this book seems to stake out a curious and unlikely middle ground,

encouraging readers to completely remodel their PC systems, equip them with Unix-based operating systems and dive into all manner of (downloaded) biological programs, but stop short at writing any serious code of their own. The true value of the book comes in laying out perhaps the first semi-comprehensive reference to the best ready-made, freely accessible bioinformatics tools available online. While other (programming) books inevitably center about reinventing the wheel, *Developing Bioinformatics Computer Skills* points the user to established and popular software applications to perform most of the key analyses in the text.

Taken together with O'Reilly's *Beginning Perl for Bioinformatics* and the ubiquitous *Learning Perl*, *Developing Bioinformatics Computer Skills* forms a good grounding and reference for the amateur bioinformatician. By itself, this title offers a light review of biological concepts, a lengthy primer on installing and using Unix (which most researchers will likely skip), and at the heart of the book, a useful reference of ready-made bioinformatics techniques, programs and databases. Still, at less than half the price of many textbooks, this very accessible reference stands alone as a cost-effective, current and basic introduction to many of the techniques that real bioinformatics workers use regularly.

Michael Seringhaus
Department of Molecular Biophysics
and Biochemistry